

Global Mapping of Microbial Functions

Katherine S. Pollard (lead) and Jonathan A. Eisen (co-lead)

Aim 1. Build automated tools to characterize OPFs from sequencing data.

Personnel

- Thomas Sharpton
- Guillaume Jospin
- Dongying Wu
- Gladstone Bioinformatics
- Katie Pollard (lead for this Aim)
- Jonathan Eisen

Tasks

Year 1

- Catalog all microbial protein families by Markov clustering and build profile HMMs (Figure 2). Compare and contrast results with other protein family databases (e.g., COGs, PFAMs).

Year 2

- Design a novel, fully automated method to classify metagenomic sequence reads into OPFs (Figure 2). Validate this OPF read classifier *in silico* by simulations.
- Implement automated methods for updating protein family alignments and HMMs as new genomes are sequenced.

Year 3

- Develop metrics to quantify OPF abundance, universality, and phylogenetic diversity from metagenomic data.
- Design bioinformatics tools for mapping known protein functions (e.g., genome annotations, domain annotations, gene ontology) onto classified metagenomic sequence reads.

Aim 2. Predict niches of OPFs from environmental data.

Personnel

- Joshua Ladau
- Gladstone Bioinformatics
- Katie Pollard (lead for this Aim)
- Jonathan Eisen

Tasks

Year 1

- Model OPF geographic distributions with statistical techniques, such as niche mapping.

Year 2

- Estimate OTU and OPF ranges in environmental niche space, defined by biogeochemical variables rather than geographic coordinates.

Year 3

- Draw geographic and environmental niche maps.

Aim 3. Analyze niche maps to characterize global functional diversity.

Personnel

- Dongying Wu
- Guillaume Jospin
- Significant input from Thomas Sharpton, Joshua Ladau, and Gladstone Bioinformatics to coordinate application of methods from Aims 1 and 2
- Jonathan Eisen (lead for this Aim)
- Katherine Pollard

Tasks

Year 1

- Assemble and process publicly available marine metagenomic data sets. Assess global sampling coverage.
- Assemble and process publicly available environmental metadata (possibly from remote data sources).

Year 2

- Apply methods from Aims 1 and 2 to marine microbial data assembled in Year 1.

Year 3

- Compare functional (OPF) and taxonomic (OTU) diversity maps to
 1. Elucidate patterns of selection, convergence, and lateral gene transfer;
 2. Develop functional hypotheses for particular gene families;
 3. Identify diversity hotspots in geographic and environmental niche space; and
 4. Propose testable hypotheses about the roles of microbial functions in ecosystem health, food webs, and biogeochemical cycles.